Deep Point Cloud Building Envelope Segmentation (DeeP-CuBES) using Deep Learning

Balaji Selvakumar, Yifang Liu, Nolan W. Hayes, Diana Hun, Bryan P. Maldonado

Buildings and Transportation Science Division, Oak Ridge National Laboratory*, United States of America

selvakumarb@ornl.gov, liuy5@ornl.gov, hayesnw@ornl.gov, hunde@ornl.gov, maldonadopbp@ornl.gov

Abstract -

Building Information Modeling (BIM) plays an important role in building design and construction, particularly for achieving energy-efficient retrofits. Building envelope retrofits using panelized prefabricated system, such as those popularized by the Energiesprong program, need accurate as-built dimensions of facade features (windows, doors, etc.) to achieve the desired thermal and air tightness. Traditionally, building surveying is done manually, resulting in a time-consuming and labor-intensive process. Recently, 3D point clouds from terrestrial LiDAR have been used to automate the generation of as-built dimensions of existing buildings. However, automated BIM using LiDAR relies on solving the point cloud semantic segmentation (PCSS) problem. In this work, we propose a robust pipeline for solving the PCSS problem using deep neural networks, focusing on overcoming challenges posed by imbalanced datasets and complex architectural features. We introduce the first high-density, labeled, and validated building envelope point cloud dataset derived from multiple building scans, specifically curated to tackle challenges in facade-level segmentation. Results from the trained neural networks show that advanced attention-based architectures and incorporating radiometry (light intensity and RGB) features significantly boost segmentation accuracy for windows and doors.

Keywords -

point cloud, semantic segmentation, supervised learning

1 Introduction

More than 40% of the buildings in the United States were constructed prior to 1980, when energy regulations were in place [1]. Building envelope retrofits using prefabricated overclad panels offer a promising solution to increase the energy efficiency of these older buildings. Moreover, this process minimizes occupant disruption and has a high potential for automation, shortening assembly time at the job site and increasing overall construction quality. However, this type of retrofit requires precise as-built dimensions of the building to effectively design and manufacture the overclad panels, highlighting the need for Building Information Modeling (BIM) to create a digital twin. Creating an as-built retrofit BIM involves measuring the geometry of building facade features, such as windows and doors, to produce a dimensional digital twin that facilitates the design and installation of prefabricated overclad panels. However, this process is often time-consuming, labor intensive, and susceptible to errors. Therefore, automating this process has substantial benefits when trying to achieve building envelope retrofits at a large scale [2].

Recent advancements in 3D LiDAR scanning technologies have enabled the acquisition of high-density point cloud data, providing detailed geometric representations of scanned structures. However, point clouds are inherently unstructured, making the extraction of semantic information—such as identifying facades, windows, and doors—a complex but essential task for creating BIM-ready datasets [3]. This process can be automated by solving the Point Cloud Semantic Segmentation (PCSS) problem, revolutionizing building surveying by eliminating the need for manual annotation and generated dimensional twins in real time. Despite its promise, several challenges remain, particularly for datasets with imbalanced class distributions and complex architectural features [4]. This study focuses on improving facade segmentation for retrofitting applications by introducing a labeled and validated high-density facade dataset, curated from multiple building scans. This dataset bridges a critical gap in resources required for evaluating advanced neural network for solving the PCSS problem.

Two deep learning models were evaluated: PointNet++, based on a multi-layer perceptron (MLP) architecture [5], and the Superpoint Transformer (SPT), a transformer-inspired model leveraging hierarchical feature learning and attention mechanisms [6]. These models were fine-tuned by training them on additional radiometry features, enabling richer feature representations and achieving superior segmentation accuracy. Furthermore, to address the

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan).

inherent challenges of imbalanced datasets, where walls dominate and classes such as windows and doors are underrepresented, we employ a combination of sampling techniques and weighted loss functions.

The structure of the paper is as follows. Section 2 reviews related work on PCSS in the construction industry, highlighting the need for labeled data. Section 3 details the dataset collection, labeling validation, post-processing procedures, and the class distribution within the dataset. Section 4 discusses the deep neural network architectures used and their evaluation metrics. Finally, Section 5 presents the results, while Section 6 presents the conclusions and future work.

2 Related Works

2.1 PCSS applications in construction

Recent breakthroughs in deep neural networks have led to the development of powerful deep learning models which can capture and learn texture, shape, and contextual information from images, accelerating the use of computer vision in various applications. Semantic segmentation is a widely used technique in computer vision, which is the process of assigning a label to every pixel in an image, indicating which category the corresponding pixel belongs to. This helps in better scene understanding of 2D images. Among various scene understanding problems, 3D semantic segmentation is helpful in determining accurate object boundaries with similar patterns along with their labels [7]. Unlike 2D images, 3D data generated by laser scanners are unordered and invariant to permutations, making it difficult for the deep neural networks to learn the desired features [3]. Due to its inherent nature, point clouds are subject to outliers and noise. In addition, non-uniform point densities in various regions affect point cloud feature learning. Deep learning (DL)-based segmentation for point clouds has been used in various applications such as robotics, industrial automation, autonomous driving [4] and 3D scene understanding, enabling automated decision making in three-dimensional space[8] . Datasets like KITTI[9] and nuScenes [10] are widely used in benchmarking point cloud segmentation algorithms and their performance in autonomous driving in real world scenarios.

Given that, in general, old buildings lack an as-built, or even an as-designed BIM, point cloud data has become popular within the construction sector for 3D BIM generation. Recently, innovative methods have been developed to efficiently create as-built BIMs for various construction projects, including envelope retrofits, re-cladding, and window replacements. [11]. PCSS can be used to generate accurate as-built BIMs from 3D laser scans of the building, where the accuracy depends mainly on the sensor and range. However, the efficiency of supervised machine learning used in PCSS is heavily dependant on the available labeled training data. Currently, the amount and variety of open source point cloud datasets to train, validate, and compare new neural network architectures are mainly limited to the autonomous driving industry. Beyond autonomous driving, numerous datasets like S3DIS featuring residential and commercial buildings focus on indoor room and office scenes captured using RGB-D devices [12]. Building exterior envelopes, however, usually correspond to sparse urban datasets like Semantic3D[13] and Toronto3D[14] collected using mobile/aerial laser scanning technology and those featuring [15] and [16] facade-level classes are of sparse density. Therefore, it is essential to address the scarcity of dense, accurately labeled point cloud data of exterior building envelopes, specifically designed for semantic segmentation tasks aimed at enhancing productivity and reducing costs in retrofit and new construction projects.

2.2 Class imbalance in real-world datasets

The performance of DL classification models is heavily impacted by how well the data is distributed across the different classes.[17] To achieve optimal performance, and avoid overfitting on individual classes, the model has to be trained on datasets that are as balanced as possible [7]. In a real-world point cloud scan of a building exterior envelope, as much as 80% of elements can be classified as walls points, while the rest correspond to features such as windows and doors.

Standard learning algorithms assume balanced class distributions and severe imbalances like our application, can lead to biased decision boundaries on under-represented minority classes. This study[18] systematically analyzed how imbalanced data skews model predictions towards majority classes, leading to biased decision boundaries. Due to the large presence of the wall points in real world facade dataset like ours, the model can be heavily skewed and can learn to over classify the majority group. Furthermore, class imbalance can distort the performance evaluation when using metrics such as accuracy, as correctly classifying the major classes may yield high scores that incorrectly indicate good performance [19].

3 Dataset Collection and Labeling

For data collection, we used the Leica MultiStation MS60 to scan buildings across the ORNL campus. The MS60 is a terrestrial laser scanner (TLS) that allows users to customize scan density, enabling the collection of dense scans with high accuracy. The lidar scanner returns the Cartesian coordinates x, y, z and the light intensity I of a point on a surface measured by the laser. In addition, the TLS is equipped with a 5 MPx RGB camera fea-

Figure 1. Sample facades used for DeeP-CuBES

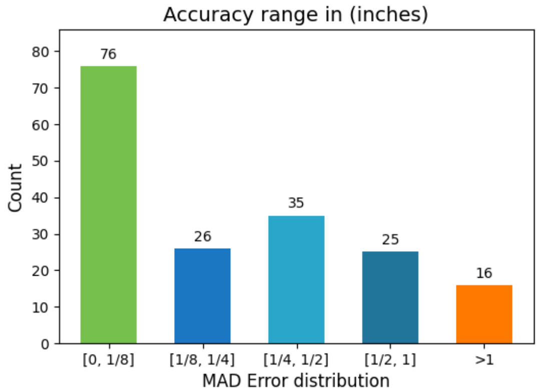


Figure 2. Error distribution for window dimensions

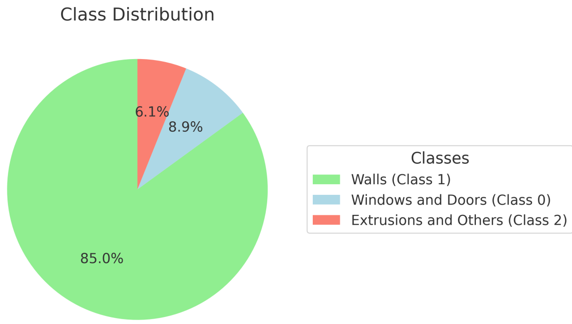


Figure 3. Distribution of 223 million points within the Oak Ridge National Laboratory Building Envelope Library (ORNOBEL)

turing a CMOS sensor which takes captures images of the scanned area to later interpolate the points with the pixel RGB values. Therefore, the final raw point cloud belongs to a seven-dimensional space with coordinates $\begin{bmatrix} x & y & z & I & r & g & b \end{bmatrix}$. A total of 25 building facades across the ORNL campus were scanned and the point clouds were initially cleaned to remove outliers such as trees and other occlusions using CloudCompare [20]. A subset of the facades collected is shown in Figure 1.

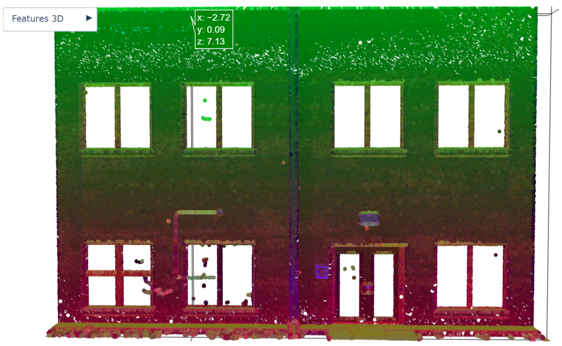
With the cleaned, pre-processed data of different individual facades, we used the ORNL-developed Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) algorithm to accelerate data labeling [21]. Auto-CuBES uses a sequence of unsupervised machine learning methods, such as k-means and DBSCAN, to segment the building envelope into wall points, windows points, door points, and other categories (building extrusions) with an accuracy of 3mm. Although Auto-CuBES is significantly faster than manual segmentation, it requires several segmentation steps which involve calibrating various parameter values to segment window and door openings from the wall points. The segmentation results from all building scans were grouped into three classes: windows+doors, walls, and others. The results from Auto-CuBES were validated by comparing them to manual measurements of the window dimensions (length and height) using the TLS laser. Figure 2 shows the distribution of Mean Absolute Deviation (MAD) observed in our dataset. Note that most of the error distribution is heavily skewed toward lower values (ranging between 0 to 1/8 in), as indicated by the high count in the first bin. Also note that the manual measurements are not the actual ground truth, as it may be subject to human inaccuracies. This result suggests that the segmentation is accurate for most cases, with less than 10% of the dataset having a MAD error greater than 1 inch. After labeling and validation, we created the Oak Ridge National Laboratory Building Envelope Library (ORNOBEL) for exterior building envelopes.

The ORNOBEL dataset consists of dense facade scans of 25 different multistory buildings with 3mm accuracy and more than 220 million points of labeled data. Figure 3 shows the distribution of the three classes used, namely windows+doors, walls, and others. These categories can be further subdivided in the future. Note that 85% of the data consists of wall points, creating a significant imbalance that can affect the performance of any trained model.

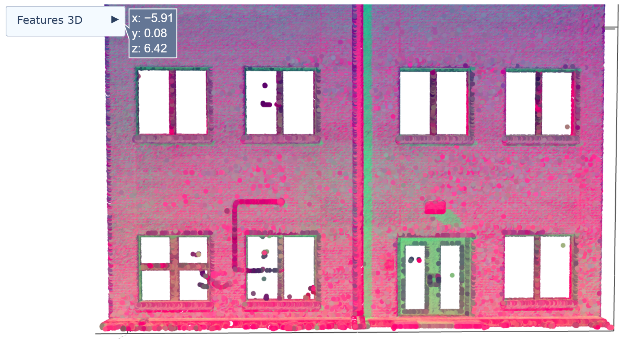
3.1 Post processing and preparing the data

To ensure the point cloud data is standardized, aligned, and ready for neural network model training, we implemented a series of pre-processing steps, as outlined below:

- **Spatial Centering & Intensity Normalization:** The origin of each facade point cloud was moved to the geometric center of its minimum bounding box. This ensures consistency of the spatial positioning of all walls and openings (windows and doors) facilitating easier model interpretation. Light intensity was normalized to avoid large values observed in the raw data.



(a) Default feature map without radiometry features



(b) Fine tuned feature map with added radiometry features

Figure 4. Comparison of fine tuned feature partition map for SPT.

To handle outliers in intensity values, we employed the *Robust Z-Score* normalization method:

$$I_{\text{robust}} = 0.6745 \cdot \frac{I - \text{Median}(I)}{\text{MAD}} \tag{1}$$

where I is the raw intensity value, $\text{Median}(I)$ is the median intensity, and MAD is the Median Absolute Deviation. This ensures that the intensity values are scaled robustly, particularly across facade surfaces and doesn't skew the model training.

- **PCA-Based Alignment:** For facade-specific data preparation, Principal Component Analysis (PCA) was applied to the wall points using scikit-learn to establish a consistent reference axis [22]. Subsequently, point clouds representing windows and doors were aligned to this PCA-derived axis to standardize their orientations. This alignment step is crucial for facade datasets collected without a common TLS orientation, as it ensures uniformity, allowing models to concentrate on geometric features such as coplanarity without being influenced by variations in orientation.

- **Output Format:** RGB color values were left unchanged, as they can serve as additional radiometric features for identifying facade components. The preprocessed point clouds, with spatial centering, normalized intensity, PCA alignment, and original RGB values, were exported in standard h5 formats. This approach ensures that the data is directly usable for neural network training.

4 Deep Neural Networks applied to PCSS

PointNet++ was a groundbreaking model introduced in 2017 to solve the Point Cloud Semantic Segmentation(PCSS) problem [5]. The algorithm pioneered the hierarchical learning of both local and global features directly from unordered point clouds. Given its historical significance, we used PointNet++ as a baseline to evaluate the PCSS accuracy on our dataset. PointNet++ uses a Multi Layer Perceptron (MLP) architecture and processes point clouds hierarchically, using set abstraction implemented by multiple MLP layers to capture local and global features. However, it only uses the xyz coordinates for feature learning, ignoring other informative attributes like light intensity and RGB. When trained and tested on the ORNOBEL dataset, the model performed poorly in capturing spatial relationships for minority classes such as windows and doors, highlighting the difficulty of dealing in imbalanced classes.

We have implemented Superpoint Transformer (SPT) neural network architecture [6], an advanced model which uses additional feature sets and super point partitions to improvise feature learning. This model draws inspiration from the U-Net convolutional architecture [23], which has been groundbreaking in 2D image segmentation. By integrating transformer-based self-attention mechanisms, which has seen breakthroughs in language understanding, SPT effectively captures the spatial dependencies between unordered points, while its U-Net-like encoder-decoder structure allows for multi-scale feature learning. Additionally, its compact size, with fewer parameters, ensures computational efficiency, making it an ideal choice for dense point cloud segmentation of building envelopes.

4.1 Superpoints features

Superpoints are groups of points clustered together based on their local geometric properties. The Superpoint Transformer outperforms PointNet++ by leveraging a more sophisticated architecture inspired by U-Net and superpoint graphs. Unlike PointNet++, SPT uses a combination of xyz coordinates, light intensity, and RGB values to construct a rich, hierarchical feature map. These features are used to construct a meaningful hierarchical partition with adjacency relationship with neighboring points. Subsequently, SPT organizes the point clouds with feature maps into superpoints (i.e., clusters of points with adaptive and shared local features) enabling efficient and structured feature aggregation.

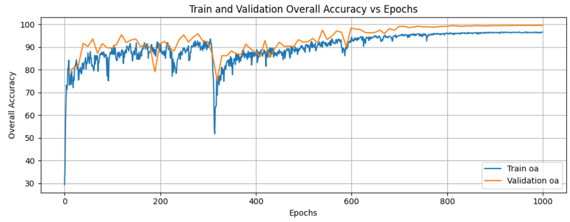
The transformer encoder-decoder architecture with self-attention allows SPT to learn spatial dependencies between points effectively, which is critical for capturing complex facade openings and features. SPT integrates MLP layers to extract point-wise features and project them onto higher dimensions before applying attention. This combines the strengths of both MLPs and transformers. Our fine-tuning process involved customizing SPT to optimize its ability to learn facade-specific features as shown in Figure 4. With the appropriate calibrated parameters in SPT, classification accuracy across all metrics was improved compared with the PointNet++ baseline results, particularly for correctly classifying minority classes.

4.2 Deep neural networks training

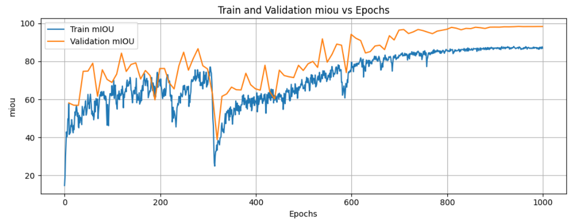
To establish a point of comparison for improvements, we generated a baseline models using PointNet++. The original PointNet++ model was trained without implementing any sampling methods or dataset balancing, relying solely on xyz coordinates for feature learning (no radiometry features like intensity or RGB were used). The model performed poorly across all classes due to the dataset's inherent imbalance. To address this, we used a combination of sampling methods: weighted cross-entropy loss as a passive sampling method and downsampling majority class as an active approach. While these techniques significantly improved the model's performance, the generalization to non-downsampled wall points remained poor. This highlighted the limitations of addressing class imbalance solely by downsampling a single class, which is impractical for real-world applications.

The Superpoint Transformer model was initially trained on the ORNOBEL dataset without fine-tuning or incorporating intensity and RGB features. Moreover, the SPT was trained directly on the class-imbalanced dataset, providing a more robust and effective solution to the challenges posed by the dataset's structure. The model performed extremely well, yielding a significant boost over the Pointnet++ model's results. To effectively train our neural networks, we employed cross-entropy loss functions. Specifically, for the PointNet++ model, we applied a weighted cross-entropy loss to explicitly handle class imbalance within the dataset. For the SuperPoint Transformer model, the standard categorical cross-entropy loss was sufficient enough.

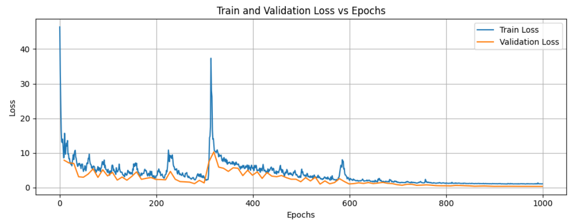
Overall Accuracy: measures the ratio of correctly classified points across the entire dataset. While accuracy was recorded, it is misleading for imbalanced datasets as it disproportionately favors the majority class (wall points in our case)



(a) Overall accuracy



(b) mIoU values



(c) Model Loss

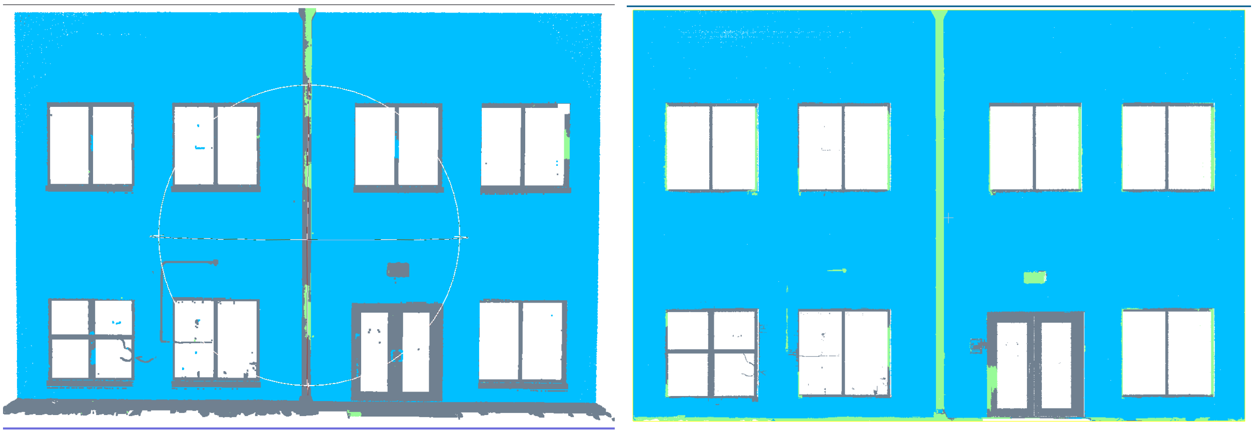
Figure 5. Comparison of training metrics from SPT

mean Intersection over Union (mIoU): this metric provides a class-wise measure of segmentation accuracy. This metric is particularly effective for semantic segmentation, as it evaluates the overlap between predicted and ground-truth points for each class.

The SPT model was trained for up to 1000 epochs with an early stopping criterion to prevent over-fitting. Pointnet++, on the other hand, was trained for 50 epochs. A standard batch size of 32 was utilized for both models, balancing computational efficiency and effective gradient updates. Figure 5 shows the three different aforementioned metrics during the training process of the SPT model for baseline comparisons. It is noteworthy that overall accuracy quickly reaches a high value >80% largely due to the bias towards wall points. Meanwhile, the mIoU values do not reach their maximum until the end of the training phase.

5 Results and Discussion

The training and testing experiments were conducted on a system equipped with an NVIDIA RTX 6000 Ada Generation GPU, Intel Xeon Gold 6442Y CPU, and 128 GB of RAM. Training times varied depending on the model complexity and number of epochs, with PointNet++ com-



(a) Pointnet++ prediction on down-sampled wall points

(b) SPT prediction on actual points

Figure 6. Comparison of model predictions.

pleting in approximately 3 hours per run and Superpoint Transformer taking around 2 hours per run. From our experiments, we observed that well-represented classes with planar geometries (such as walls) are correctly segmented in most of the instances. Figure 6b shows, the segmentation results for the Pointnet++ (left) and SPT (right). The improved segmentation accuracy of SPT model resulted from careful feature fine-tuning, where additional point features (such as RGB, intensity, and elevation) were integrated beyond the default set used in the original implementation.

Note that, in general, the wall points (denoted in blue color) are well identified in both models. However, opening points (grey color) were more accurately segmented using SPT. In the case of Pointnet++, such points were not distinctly separated from the third class (green color)which corresponded to other protrusions, such as gutters or ground platforms. The testing error metrics shown in Table 1 demonstrates that the SPT has a superior performance for solving the PCSS problem for building envelopes. This improved performance is primarily attributed to the SPT's attention-based architecture, which effectively captures detailed spatial relationships and long-range contextual features. Additionally, incorporating radiometric information, such as RGB and intensity values, enables richer feature representations, significantly improving segmentation accuracy for facade elements like windows and doors. It is important to note, however, that RGB features do not inherently resolve class imbalance; instead, the enhanced accuracy from radiometric features improve overall model performance.

After the PCSS problem is solved, the digital twin of the building envelope is created by finding the minimum bounding box of the different classes. Figure 7 shows the bounding box for each opening. Note that the SPT model struggled with accurate predictions along the boundaries of windows and doors, where the transition between classes introduces ambiguities leading to partition errors that cannot be fully corrected with the current model. This boundary defines the transition between classes and lacks a distinct geometric or feature discontinuities where two or more classes intersect. Table 2 compares the dimensions of the digital twin extracted using the SPT model versus the original one defined using the Auto-CuBES algorithm. The opening are number from 1 to 8, going from left to right, and from bottom to top. Note that the bottom openings show the largest errors, while the top openings show a low height error and large width error. Upon close inspection, the errors are mainly due to outliers in the opening classes. One of the last steps in Auto-CuBES involve removing outliers from each identified cluster, drastically increasing its accuracy. The SPT model, on the other hand, its very efficient at identifying the main points belonging to a particular cluster, but cannot completely remove the outliers within a cluster. This exemplifies the limitations of relying solely on a deep learning model, and directs our research towards a more robust PCSS solution that uses both neural networks and unsupervised clustering techniques.

Table 1. Testing error metrics on ORNOBEL Dataset

Metrics	mIOU	OA%
Pointnet++ (downsampled)	93	97%
Superpoint Transformer	97.3	98.7%

6 Conclusion and Future Work

In this work, we present ORNOBEL, a high density, labeled 3D facade segmentation dataset specifically de-

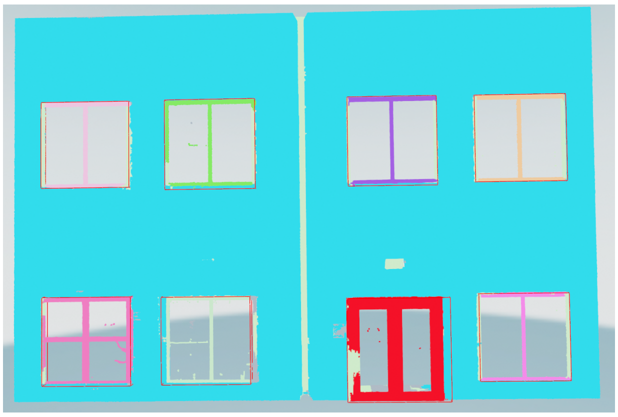


Figure 7. Bounding box applied to SPT results

Table 2. Comparison between digital twin dimensions generated by SPT and Auto-CuBES

Open	SPT (m)		Auto-CuBES		Error (mm)	
	Height	Width	Height	Width	Hght	Wdth
1	2.520	2.211	2.186	1.947	334	264
2	2.141	2.134	1.945	1.864	196	27
3	2.135	1.879	1.943	1.838	192	41
4	1.924	1.866	1.943	1.835	-19	31
5	1.968	1.954	1.956	1.856	12	98
6	1.946	1.940	1.945	1.854	1	86
7	1.960	1.886	1.944	1.843	16	43
8	1.964	1.886	1.952	1.854	16	43

signed for building exterior envelopes, addressing the lack of labeled datasets for building retrofitting applications. Our findings highlight the potential of 3D PCSS in generating precise Building Information Models (BIMs) for retrofitting applications, particularly for designing and deploying prefabricated overclad panels to enhance building energy efficiency. Based on the observations from 2D image-based facade segmentation domain [24, 2], we believe that our 3D facade segmentation dataset will contribute to the further development of 3D facade-oriented methods for construction domain. Results from the deep learning models tested show that a transformer-based architecture has a good potential for solving the semantic segmentation problem in dense building envelope point clouds. However, we plan to extend our work by further training on more state-of-the-art models in combination with various sampling methods and clustering techniques to further address our current limitations. For future work, we aim to expand our dataset by including additional building scans with diverse architectural styles and materials, removing PCA to generalize and address class boundary limitations. Additionally, we will open-source our dataset to address the lack of labeled resources in this domain, fostering collaboration and innovation in 3D facade segmentation. Future exploration will also involve developing methods for faster and improved real-time segmentation capabilities for on-site retrofitting applications through accelerated pipelines.

Acknowledgements

This research was supported by the DOE Office of Energy Efficiency and Renewable Energy (EERE), Building Technologies Office (BTO), under the guidance of Sven Mumme, and used resources at the Building Technologies Research and Integration Center, a DOE-EERE User Facility at Oak Ridge National Laboratory (ORNL). This research was also sponsored by the Laboratory Directed Research and Development Program of ORNL, managed by UT-Battelle, LLC, for the U. S. Department of Energy

References

[1] Iea. Renovation of near 20% of existing building stock to zero-carbon-ready by 2030 is ambitious but necessary – analysis, Sep 2022.

[2] Menglin Dai, Wil O.C. Ward, Gregory Meyers, Danielle Densley Tingley, and Martin Mayfield. Residential building facade segmentation in the urban environment. Building and Environment, 199:107921, 2021. ISSN 0360-1323. doi:https://doi.org/10.1016/j.buildenv.2021.107921. URL `https://www.sciencedirect.com/science/article/pii/S0360132321003243`.

[3] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation, 2017. URL `https://arxiv.org/abs/1612.00593`.

[4] Lukas Rauch and Thomas Braml. Semantic point cloud segmentation with deep-learning-based approaches for the construction industry: A survey. Applied Sciences, 13(16), 2023. ISSN 2076-3417. doi:10.3390/app13169146. URL `https://www.mdpi.com/2076-3417/13/16/9146`.

[5] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413, 2017.

[6] Damien Robert, Hugo Raguet, and Loic Landrieu. Efficient 3d semantic segmentation with superpoint transformer, 2023. URL `https://arxiv.org/abs/2306.08045`.

[7] Anh Nguyen and Bac Le. 3d point cloud segmentation: A survey. In 2013 6th

IEEE Conference on Robotics, Automation and Mechatronics (RAM), pages 225–230, 2013. doi:10.1109/RAM.2013.6758588.

[8] Chao Yin, Boyu Wang, Vincent J.L. Gan, Mingzhu Wang, and Jack C.P. Cheng. Automated semantic segmentation of industrial point clouds using respointnet++. Automation in Construction, 130:103874, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.103874. URL https://www.sciencedirect.com/science/article/pii/S0926580521003253.

[9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[10] Whye Kit Fong et al. Panoptic nuscenes: A large-scale benchmark for lidar panoptic segmentation and tracking. 2021.

[11] Xuehan Xiong, Antonio Adan, Burcu Akinci, and Daniel Huber. Automatic creation of semantically rich 3d building models from laser scanner data. Automation in Construction, 31:325–337, 2013. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2012.10.006. URL https://www.sciencedirect.com/science/article/pii/S0926580512001732.

[12] O. Wysocki, L. Hoegner, and U. Stilla. Tumfacade: Reviewing and enriching point cloud benchmarks for facade segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-2/W1-2022: 529–536, 2022. doi:10.5194/isprs-archives-XLVI-2-W1-2022-529-2022.

[13] Timo Hackel, N. Savinov, L. Ladicky, Jan D. Wegner, K. Schindler, and M. Pollefeys. SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, volume IV-1-W1, pages 91–98, 2017.

[14] Weikai Tan, Nannan Qin, Lingfei Ma, Ying Li, Jing Du, Guorong Cai, Ke Yang, and Jonathan Li. Toronto-3D: A large-scale mobile lidar dataset for semantic segmentation of urban roadways. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 202–203, 2020.

[15] F. Matrone, A. Lingua, R. Pierdicca, E. S. Malinverni, M. Paolanti, E. Grilli, F. Remondino, A. Murtiyoso, and T. Landes. A benchmark for large-scale heritage point cloud semantic segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2020:1419–1426, 2020.

[16] Olaf Wysocki, Jiarui Zhang, and Uwe Stilla. TumfaÇade, 2021. URL https://mediatum.ub.tum.de/1636761.

[17] Kushankur Ghosh, Colin Bellinger, Roberto Corizzo, Paula Branco, Bartosz Krawczyk, and Nathalie Japkowicz. The class imbalance problem in deep learning. Machine Learning, 113 (7):4845–4901, Jul 2024. ISSN 1573-0565. doi:10.1007/s10994-022-06268-8. URL https://doi.org/10.1007/s10994-022-06268-8.

[18] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9):1263–1284, 2009. doi:10.1109/TKDE.2008.239.

[19] Justin M. Johnson and Taghi M. Khoshgoftaar. Survey on deep learning with class imbalance. Journal of Big Data, 6(1), Mar 2019. doi:10.1186/s40537-019-0192-5.

[20] CloudCompare Contributors. CloudCompare (version 2.13.2). http://www.cloudcompare.org/, 2024. GPL software.

[21] Bryan P. Maldonado, Nolan W. Hayes, and Diana Hun. Automatic point cloud building envelope segmentation (auto-cubes) using machine learning. Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC), Jul 2023. doi:10.22260/isarc2023/0009.

[22] F. et al. Pedregosa. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. CoRR, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.

[24] Nariman Habili, Ernest Kwan, Weihao Li, Christfried Webers, Jeremy Oorloff, Mohammad Ali Armin, and Lars Petersson. A hyperspectral and rgb dataset for building facade segmentation, 2022. URL https://arxiv.org/abs/2212.02749.